Konnexions ML

Report on Final Project – Spam Detection

Submitted by

<u>Mayank Gajwe</u>

Roll: 21052080

Computer Science Engineering, KIIT

Under the Guidance of

Gyanig K. & Shivank Singh

In this project we will classify mails into ham and spam mails using machine learning algorithm with the assumption that words are independent of each other using Naïve Bayes Algorithm. This project uses the Multinomial Naïve Bayes algorithm from the SciKit Learn Library.

Naïve Bayes classification is a simple probability algorithm based on the fact, that all features of the model are independent. In the context of the spam filter, we suppose, that every word in the message is independent of all other words and we count them with the ignorance of the context. The text if filtered with the stopwords library so that words such as has, the, it, etc. will be ignored as they do not serve any purpose in the classification of text.
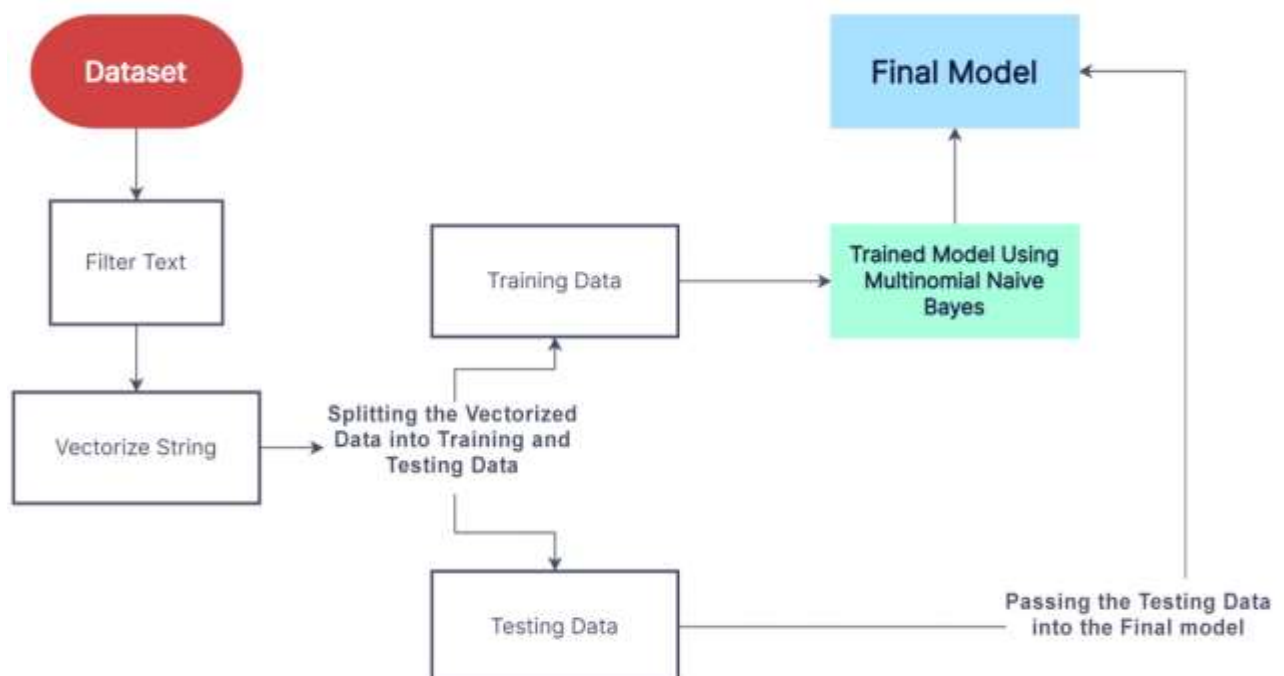
Final Model:

```
[0 0 1 ... 1 0 1]
[0 0 1 ... 0 0 1]
              precision      recall   f1-score    support

           0       1.00        0.99       0.99        885
           1       0.95        0.99       0.97        261

    accuracy                              0.99       1146
   macro avg       0.97        0.99       0.98       1146
weighted avg       0.99        0.99       0.99       1146


Accuracy for testing data:  0.9869109947643979
```